2.4 HIGHWAY CAPACITY

2.4.1 General Characteristics

The term "capacity" is used to express the maximum hourly rate at which persons or vehicles can reasonably be expected to traverse a point (i.e., a uniform section of a lane or a roadway) during a given time period under prevailing roadway and traffic conditions. The range of traffic flow on a highway can vary from very light volumes to volumes that equal the capacity of the facility as defined above. In the generic sense, the term also encompasses broader relations between highway characteristics and conditions, traffic composition and flow patterns, and the relative degree of congestion at various traffic volumes. Highway capacity issues in this broad sense are discussed below.

Sections 2.4.2 through 2.4.6 provide a brief overview of the principles and major factors concerning highway design capacity. To determine the capacity for a particular highway design, the designer should refer to the *Highway Capacity Manual* (HCM) (*37*) for guidance. The HCM is used as the basic reference for the following discussion.

2.4.2 Application

Highway capacity analysis serves three general purposes, including:

- **Transportation planning studies**—Highway capacity analysis is used in these studies to assess the adequacy or sufficiency of existing highway networks to service current traffic. In addition, it is used to estimate the time in the future when traffic growth may exceed the capacity of a highway or perhaps reach a level of congestion below capacity that is considered undesirable.

- **Highway design**—A knowledge of highway capacity is essential to properly fit a planned highway to traffic demands. Highway capacity analysis is used both to select the highway type and to determine dimensions such as the number of lanes and the minimum lengths for weaving sections.

- **Traffic operational analyses**—Highway capacity analysis is used in these analyses for many purposes, but especially for identifying bottleneck locations (either existing or potential). It is also used to estimate operational improvements that may result from prospective traffic control measures or from spot alterations in the highway geometry.

The traffic data for these uses varies with the degree of accuracy needed. For traffic-operational analyses, in which the success of minor improvements may be measured in terms of a few vehicles per hour, a high degree of precision is desirable. For highway design, a much lower order of precision suffices because the traffic data are frequently estimated for a period 10 to 20 years in the future and involve not only approximations of traffic volumes but also approximations of such factors as traffic composition and movement patterns. The discussion below shows the appropriate level of detail to achieve a reasonable balance between the design of the highway and the estimated future traffic. Such an analysis should verify that future operating conditions will not fall below an acceptable level. If a greater accuracy than is available from the suggested procedures is needed, refer to the HCM (*37*) and other reports on traffic operational analysis.

2.4.3 Capacity as a Design Control

Design Service Flow Rate versus Design Volume

The design volume is the volume of traffic projected to use a particular facility during the design year, which is usually 10 to 20 years in the future. Design volumes are estimated in the planning process and are often expressed as the expected traffic volume during a specified design hour. The derivation of the DHV has been discussed in Section 2.3, "Traffic Characteristics."

Design service flow rate is the maximum hourly flow rate of traffic that a highway with particular design features would be able to serve without the degree of congestion falling below a pre-selected level, as described in "Acceptable Degrees of Congestion."

A major objective in designing a highway is to create a facility with dimensions and alignment that can serve the design service flow rate, which should be at least as great as the flow rate during the peak 15-minute period of the design hour, but not so great as to represent an extravagance in the design. Where this objective is accomplished, a well-balanced, economical highway facility will result.

Measures of Congestion

Three key considerations in geometric design are the roadway design, the traffic using the roadway, and the degree of congestion on the roadway. The first two considerations can be measured in exact units. For example, the roadway either is or is not a highway with full control of access, its cross-section dimensions can be expressed in meters [feet], and the steepnesses of its grades can be expressed as a percentage. Likewise, traffic flow can be expressed as the number of vehicles per unit of time, traffic composition can be expressed as the percentage of vehicles of each class, and the peaking characteristics and directional distribution of traffic can also be quantified.

A scale of values for expressing the degree of congestion is, however, a much more elusive measure. Numerous measures of the overall service provided by a roadway section have been suggested, including crash frequency and severity, freedom to maneuver, the ratio of traffic volume to capacity (v/c), operating speed, average running speed, and others. In the case of signalized intersections, the stopped delay encountered by motorists is a commonly used measure of congestion.

For uninterrupted traffic flow (i.e., flow not influenced by signalized intersections), traffic operational conditions are defined by using three primary measures: speed, volume (or rate of flow), and density. Density describes the proximity of vehicles to one another and reflects the freedom to maneuver within the traffic stream. It is a critical parameter describing traffic operations with uninterrupted flow. As density increases from zero, the rate of flow also increases because more vehicles are on the roadway. While this is happening, speed begins to decline (due to the vehicle interactions). This decline is virtually negligible at low densities and flow rates. However, as density continues to increase, a point is reached at which speed declines noticeably. A maximum rate of flow is eventually reached at which the high density of traffic results in markedly decreased speeds and a reduced flow rate. This maximum rate of flow for any given facility is defined as its capacity. As capacity is approached, flow becomes more unstable because available gaps in the traffic stream become fewer and fewer. At capacity, there are no usable gaps in the traffic stream, and any conflict from vehicles entering or leaving the facility, or from internal lane changing maneuvers, creates a disturbance that cannot be effectively damped or dissipated. Thus, operation at or near capacity is difficult to maintain for long periods of time without the formation of upstream queues, and forced or breakdown flow becomes almost unavoidable. For this reason, most facilities are designed to operate at volumes less than their capacity.

For interrupted flow, such as that occurring on streets where traffic is controlled by signals, the highway user is not as concerned with attaining a high travel speed as with avoiding lengthy stops at intersections or a succession of stops at several intersections. Average stopped-time delay is the principal measure of effectiveness used in evaluating signalized intersections. Stopped-time delay, which is used because it is reasonably easy to measure and is conceptually simple, is a characteristic of intersection operations that is closely related to motorist perceptions of quality of traffic flow.

Relation between Congestion and Traffic Flow Rate

Congestion does not necessarily involve a complete stoppage of traffic flow. Rather it can be thought of as a restriction or interference to normal free flow. For any given class of highway, congestion increases with an increase in flow rate until the flow rate is almost equal to the facility's capacity, at which point congestion becomes acute. The gradual increase in congestion with increase in flow rate is apparent no matter what measure is used as an index of congestion.

The relationship between running speed and traffic flow rate for freeways, multilane highways, and two-lane highways has been discussed earlier in Section 2.3.6 on "Speed." As the traffic flow rate approaches a facility's capacity, as defined in the HCM (*37*), any minor disruption in the free flow of traffic may cause traffic on a roadway to operate on a stop-and-go basis, with a resulting decrease in traffic flow rate that can be served.

Highway sections where the paths of traffic merge and diverge within relatively short distances are called "weaving sections." Average running speed, and hence the degree of congestion, is a function not only of the volume of traffic involved in the weaving (crossing) movements but also of the distance within which the weaving maneuvers are completed. (Weaving is addressed in Section 2.4.6 under "Weaving Sections.")

On arterial streets within the urban environment, average running speed varies only slightly with changes in traffic flow rate. However, delay at signalized intersections may increase dramatically as flow rates approach capacity. Therefore, greater degrees of congestion occur, and these result in reduced overall travel speeds, higher average travel times, and traffic spill-backs into upstream intersections.

Acceptable Degrees of Congestion

From the standpoint of the highway user, it would be preferable for each user to have an exclusive right to the highway at the time the motorist finds occasion or need to use it. Moreover, a motorist would prefer that all highways be of types that would permit speeds far in excess of those normally afforded by urban surface streets. However, users recognize that if others are to share in the costs of transportation facilities, they are also entitled to share in their use. Therefore, they will readily accept a moderate amount of congestion. Just what degree of congestion the motoring public is willing to accept as reasonable remains a matter of conjecture, but it is known to vary with a number of factors.

The average motorist understands in a general sense that corrective measures to alleviate congestion may be more costly in some instances than in others. As a result, motorists will generally accept a higher degree of congestion in those areas where improvements can be made only at a substantial cost. Also, motorists are more willing to accept a higher degree of restraint in short trips than they are in long trips, but motorists are generally not satisfied with the type of operation that occurs when the volume of traffic approaches the facility's capacity.

From a highway administrator's point of view, the degree of congestion that highway users experience is related to the availability of resources. Historically, funds have never been sufficient to meet all needs, causing severe strain in improving highways rapidly enough to prevent the traffic demand from exceeding the capacity of the facility.

The appropriate degree of congestion that should be used in planning and designing highway improvements is determined by weighing the desires of the motorists against the resources available for satisfying these desires. The degree of congestion that should not be exceeded during the design year on a proposed highway can be realistically assessed by: (1) determining the operating conditions that the majority of motorists will accept as satisfactory, (2) determining the most extensive highway improvement that the governmental jurisdiction considers practical, and (3) reconciling the demands of the motorist and the general public with the finances available to meet those demands.

This reconciliation of desires with available resources is an administrative process of high importance. The decision should first be made as to the degree of congestion that should not be exceeded during the design period.

2.4.4 Factors Other Than Traffic Volume That Affect Operating Conditions

The ability of a highway to serve traffic efficiently and effectively is influenced by the characteristics of the traffic and by the design features of the highway.

Highway Factors

Few highways have ideal designs. Although most modern freeways have adequate cross-sectional dimensions, many are not ideal with respect to design speed, weaving section design, and ramp terminal design. Inadequacies in these features will result in inefficient use of the remaining portions of the freeway.

On other classes of multilane highways, intersections, even though unsignalized, often interfere with the free-flow operation of traffic. Development adjacent to the highway with attendant driveways and interference from traffic entering and leaving the through-traffic lanes cause an increase in congestion and may increase crash frequency even at relatively low volumes. The adverse effect, although readily apparent, can be difficult to quantify (*13*). Sharp curves and steep grades cannot always be avoided, and it is sometimes appropriate to compromise on cross-sectional dimensions. All of these conditions combine to cause congestion to be perceived at lower traffic volumes than would be the case for highways designed with ideal features and protected by full access control or by access management.

For urban streets with signalized intersections at relatively close intervals, the traffic volumes that could otherwise be served are reduced because a portion of each signal cycle is assigned exclusively to the crossing highway.

For a highway that is deficient in some of its characteristics and where the traffic stream is composed of a mixture of vehicle classes rather than passenger cars only, compensatory adjustment factors need to be applied to the traffic flow rates used as design values for ideal highway conditions. These adjustments are